

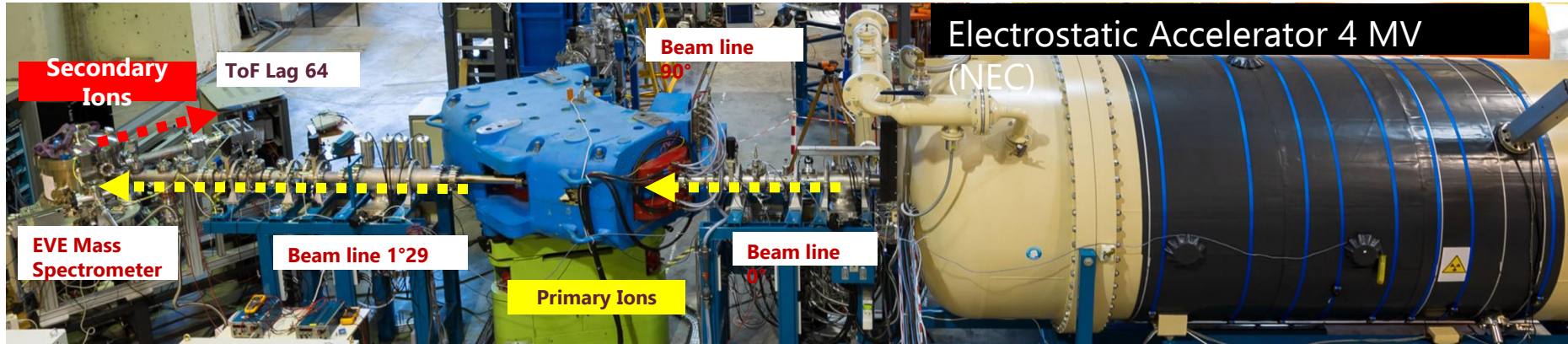
Projet R&T AIMS: Artificial Intelligence for Mass Spectrometry

IJCLab Orsay (laboratoire porteur): Françoise Bouvet, Serge Della Negra, Isabelle Ribaud, **Thomas Zerguerras**

LPSC Grenoble: Julien Angot

Contact : thomas.zerguerras@ijclab.in2p3.fr

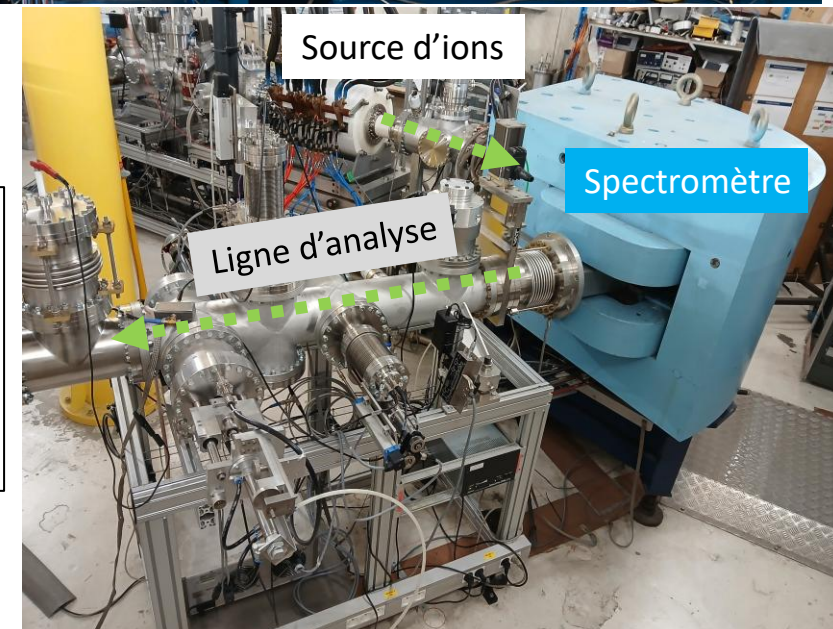
Plateforme MOSAIC Andromède (ANR: 10-EQPX-0023) @ IJCLab : SIMS-TOF



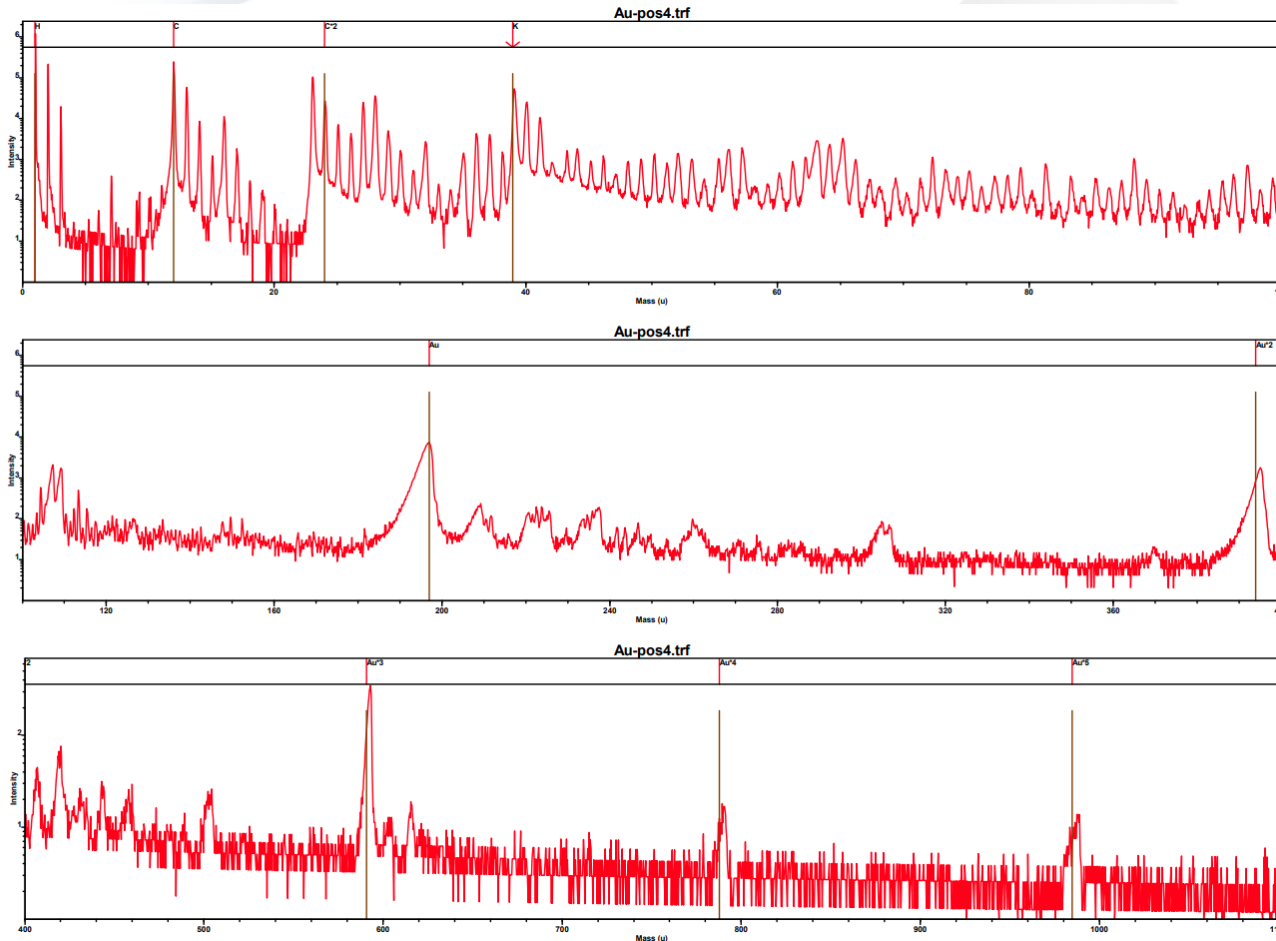
Source d'ions:
LMIS (AuSi)

Ligne 1+-N+: diagnostic pureté de sources pour accélérateurs

- Source d'ions (Booster): ECR, 14.5GHz, 100-800W, HT: 20kV
- Spectromètre: 120°, rayon 0.75m
- Mesure de courant:
 - Coupelle de Faraday avec anneau électrostatique
 - Fentes et channeltron



Applications: physique nucléaire, astrophysique nucléaire, astrochimie, science des matériaux, biologie ...



- Identifier les pics de manière **rapide** et **automatique**
- **Développer des modèles prédictifs d'attribution des pics :**
 - rapport isotopique, reconnaissance de motif ou famille de pics
 - fragmentation, clusters
 - autres contaminants, substrat

Difficultés :

- Obtention de données homogènes depuis des sources diverses
- Complexité des spectres de masse (plusieurs centaines de pics par spectre)
- Collecte et construction de métadonnées, standardisation
- Volume limité de données expérimentales

Données expérimentales

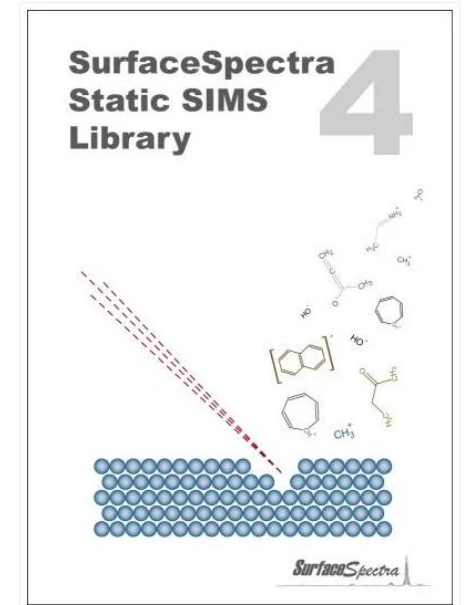
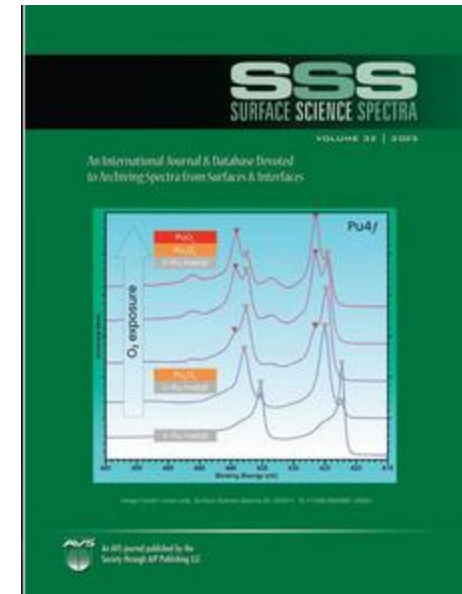
- MOSAIC/ANDROMEDE
- LPSC Grenoble
- GANIL Caen (discussion en cours)

Données bibliographiques

- Surface Science Spectra

Base de données constructeurs

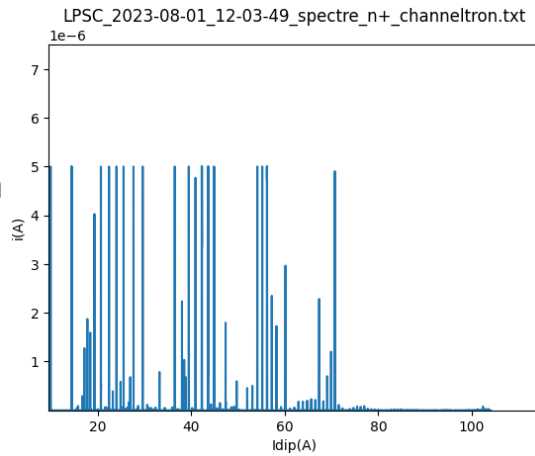
- Surface Spectra Static SIMS Library





Pré-traitement des données

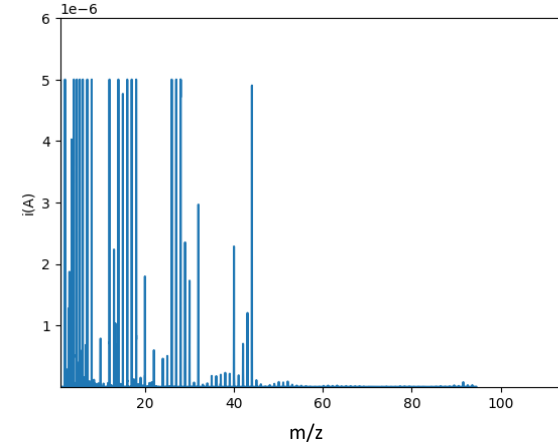
Calibration
 m/z



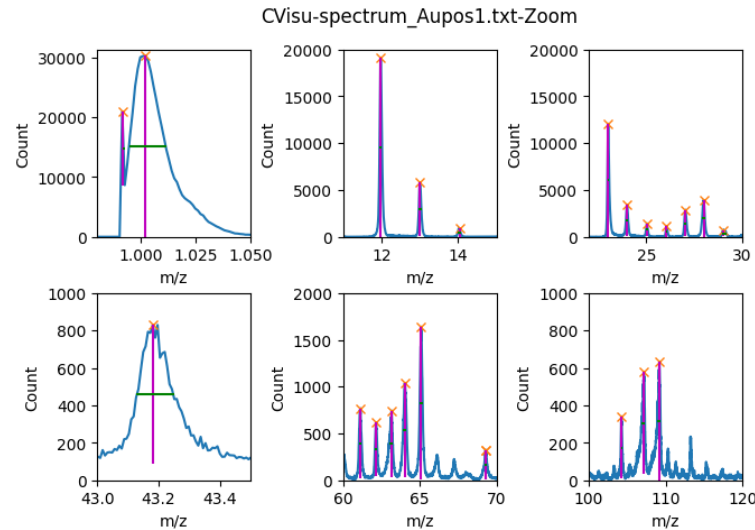
Fichier « Tags »

```
A;q;X;i;m;Species;Remark;Interpolation
12;1;36.488414E+0;4.999886E-6;12.000000;C;
12;2;25.536465E+0;5.000000E-6;12.000000;C;
12;3;20.690348E+0;5.003642E-6;12.000000;C;
12;4;17.792888E+0;1.864848E-6;12.000000;C;
12;5;15.818297E+0;89.316278E-9;12.000000;C;
14;1;39.479189E+0;4.999964E-6;14.00307481;N;
14;3;22.419327E+0;5.000000E-6;14.00307481;N;
14;4;19.204212E+0;4.016665E-6;14.00307481;N;
14;5;17.161425E+0;1.277470E-6;14.00307481;N;
16;2;29.619253E+0;5.000000E-6;15.99491464;O;
16;3;24.020579E+0;5.000000E-6;15.99491464;O;
16;5;18.403417E+0;1.588881E-6;15.99491464;O;
16;6;16.720636E+0;291.187600E-9;15.99491464;O;
40;1;67.340118E+0;2.282377E-6;39.9623831;Ar;
40;2;47.225338E+0;1.000001E-6;39.9623831;Ar;
40;3;38.482889E+0;1.833991E-6;39.9623831;Ar;
40;4;33.287861E+0;783.622106E-9;39.9623831;Ar;
40;6;26.957070E+0;675.644402E-9;39.9623831;Ar;
40;7;24.084077E+0;584.381302E-9;39.9623831;Ar;
40;8;23.220814E+0;390.000400E-9;39.9623831;Ar;
40;9;21.042072E+0;61.867176E-9;39.9623831;Ar;
```

LPSC_2023-08-01_12-03-49_spectre_n+_channeltron.txt A/q calibrated



Recherche automatique et caractérisation de pics

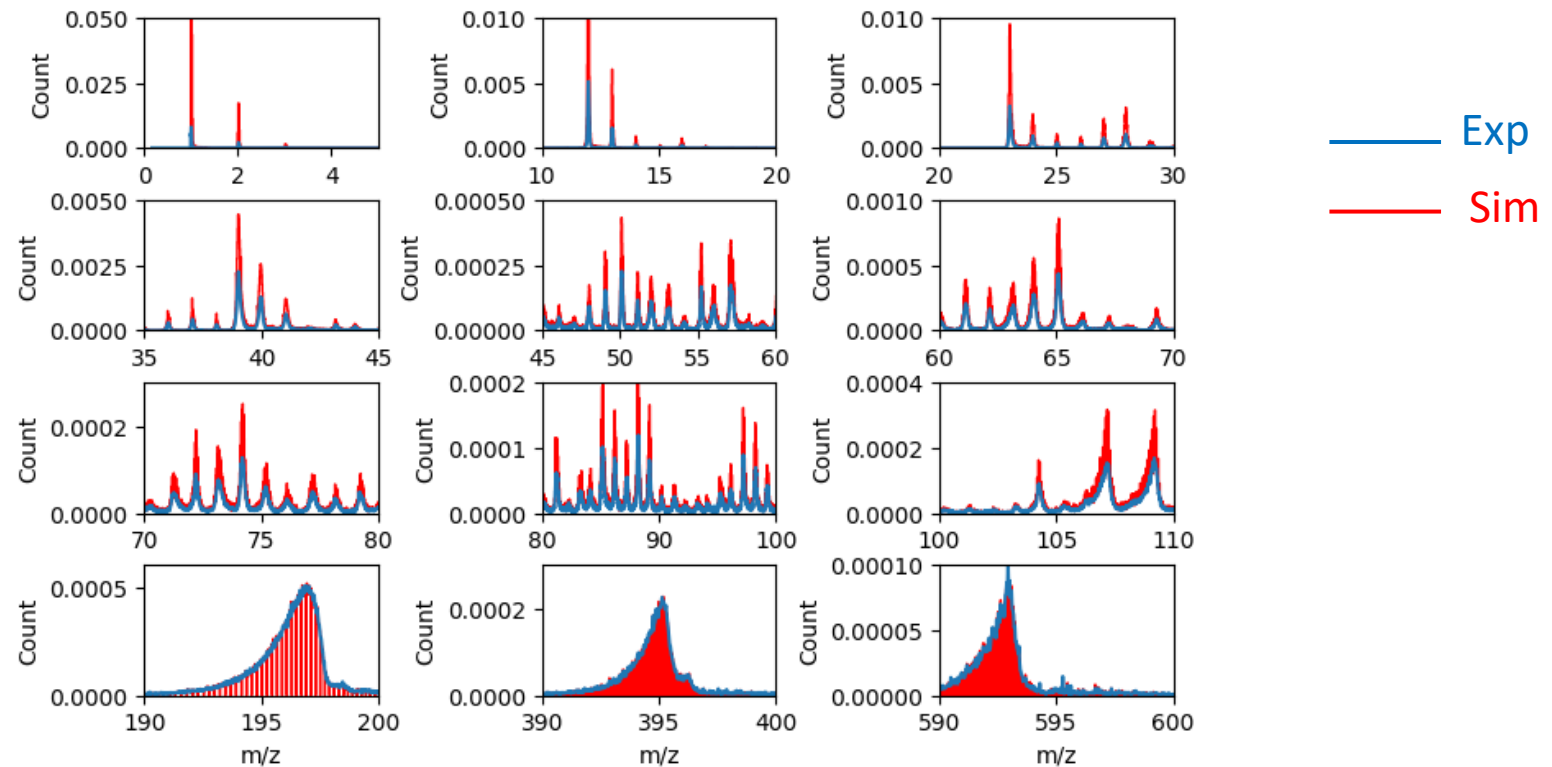




Simulations

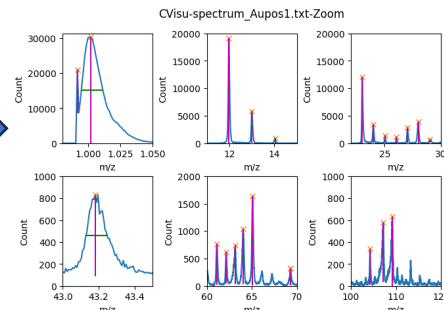
- Objectif: augmenter la quantité de données
- Génération à partir de spectres expérimentaux (normalisation -> densité de probabilité)
- Intégration progressive de nouveaux échantillons de référence et introduction de perturbations (bruit, shift ...)

Travail en cours





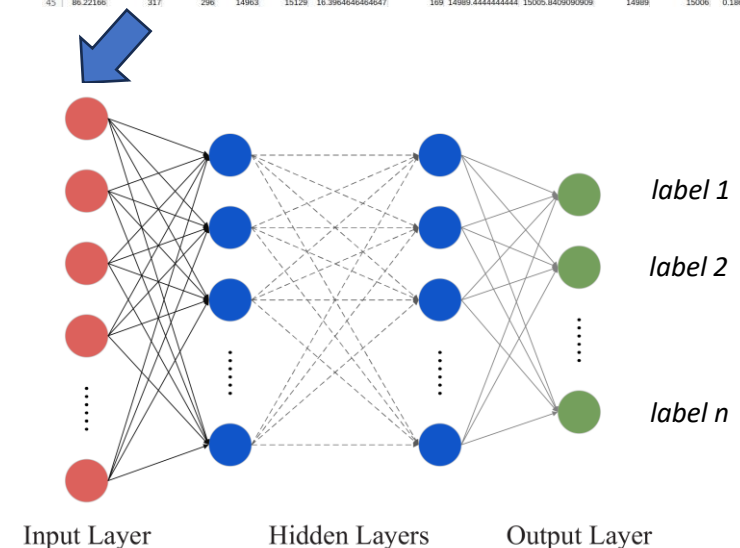
Données (expérimentales, simulations)



Données spectres

Données pour entraînement et validation

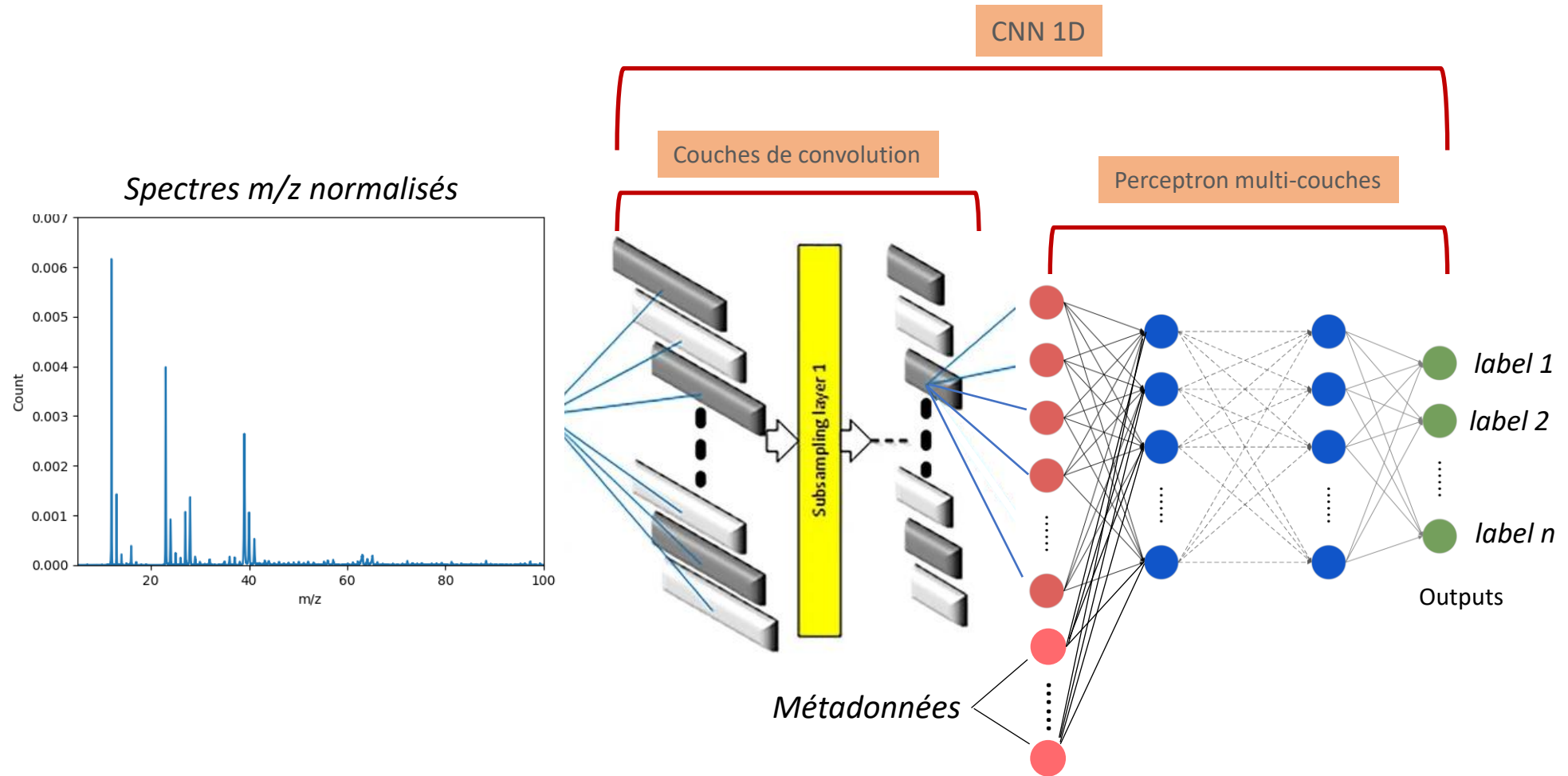
	A	B	C	D	E	F	G	H	I	J	K	L	M	N					
	area	begin	height	prominences	left	base	right	base	width	width	height	left	right	width	ent	width	ent	width	ent
1	2.09142	20095	12789	969	1001	17.93760177439483	14703	96	1000	1000	1001	1001.0000000000000000	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
2	1.30239	1989	1399	1389	1484	10040111614	17300	1389	1389	1389	1471.1338619441	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
3	2.01235	8051	18051	1680	2038	13.562605113826	4025	1707	9164	55623423	1721.471721472	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
4	3.10199	1809	989	2238	127	13.456647108617	494	5	247	38	122.256617801	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
5	11.94545	19125	16797	1927	1925	16.9072222222222222	969	1455	5556	5556	1512.5156	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
6	13.0058	5800	5741	5352	5756	15.004100728536	2924	5	540	1326	1041.06785	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
7	10.03837	901	917	5502	5756	15.0243939394	470	5	540	1326	1041.06785	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
8	10.03837	901	917	5502	5756	15.0243939394	470	5	540	1326	1041.06785	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
9	10.03837	12078	12078	7329	8972	16.07210880133	6039	1471	1651	9047	1742.50627939	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
10	11.239138	3414	3282	7541	8018	18.289724470031	1773	759	892	8982	7594.06119505	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
11	23.9138	1773	1773	1773	1773	16.512746179099	1773	1773	1773	1773	1773.1773	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
12	26.04688	1137	1178	7856	8018	17.927246487059	759	7823	6292	6292	7464.621390505	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
13	47.270543	2884	2958	8047	8168	20.071781958423	1456	807	9437	7310	8017.6117480704	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
14	27.01371	1773	1773	1773	1773	16.512746179099	1773	1773	1773	1773	1773.1773	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
15	29.20131	736	711	8349	8972	32.82470582853	3853	8402	20666666666666666666	834.913721254002	8082	8435	0.21067	0.0015000000000000000000	0.0004000000000000000000	0.0004000000000000000000	0.0004000000000000000000		
16	36.01716	1134	1129	9305	9529	18.8307768074186	571	9433	8430	7387.0067742	8444	9462	0.1279799999999999999999	0.0071100000000000000000	0.000100000000000000000000	0.000100000000000000000000	0.000100000000000000000000		
17	37.04872	1652	1652	1652	1652	16.5062222222222222	1652	1652	1652	1652	1652.1652	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
18	38.0865	968	968	968	978	17.007407485225	433	9732	20900000000000000000	9747.8320695149	9742	9748	0.1169799999999999999999	0.000100000000000000000000	0.000100000000000000000000	0.000100000000000000000000	0.000100000000000000000000		
19	39.01356	8482	8482	9385	16420	21.313401346841	4241	9857	142310000000000000000	9878.6411693735	9857	9878	0.1541000000000000000000	0.000739999999999999999999	0.007410000000000000000000	0.007410000000000000000000	0.007410000000000000000000		
20	39.97382	4800	4562	9951	16420	21.313401346841	2519	9981	1625	1005.2996350789	9983	1005	0.1541000000000000000000	0.000739999999999999999999	0.007410000000000000000000	0.007410000000000000000000	0.007410000000000000000000		
21	40.01356	18420	18420	18420	18420	18.4200000000000000	18420	18420	18420	18420	18420.18420	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
22	43.1112	830	744	10376	10978	15.181481481481	498	10407	35555555555555555555	10422.5307307307	10407	10423	0.1245499999999999999999	0.000773999999999999999999	0.000773999999999999999999	0.000773999999999999999999	0.000773999999999999999999		
23	44.0818	497	419	10473	10978	15.097560179817	267	5	10510	15	10511	10510.7650976	0.00117999999999999999	0.000786000000000000000000	0.000786000000000000000000	0.000786000000000000000000	0.000786000000000000000000		
24	45.0828	10978	11097	11651	11914	17.919141914191	1099	11651	11914	11914	11914.11914	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	0.00117999999999999999	
25	46.06678	574	564	10978	11230	17.229062323152	597	11139	86011111111111111111	11156.590162944	11139	11157	0.1493800000000000000000	0.008299999999999999999999	0.008299999999999999999999	0.008299999999999999999999	0.008299999999999999999999		
26	50.11781	854	840	10978	11789	15.843174032957	947	11184	896551741	11280.779776214	11189	11281	0.1341799999999999999999	0.008380000000000000000000	0.008380000000000000000000	0.008380000000000000000000	0.008380000000000000000000		
27	50.11781	854	840	10978	11789	15.843174032957	947	11184	896551741	11280.779776214	11189	11281	0.1341799999999999999999	0.008380000000000000000000	0.008380000000000000000000	0.008380000000000000000000	0.008380000000000000000000		
28	52.00526	427	404	11454	11789	16.225	11454	11789	11789	11789.11789	11454	11510	0.22213	0.008300000000000000000000	0.008599999999999999999999	0.008599999999999999999999	0.008599999999999999999999		
29	53.14684	334	321	11578	11789	16.225	11578	11789	11789	11789.11578	11578	11649	0.2462999999999999999999	0.008620000000000000000000	0.008620000000000000000000	0.008620000000000000000000	0.008620000000000000000000		
30	53.14684	334	321	11578	11789	16.225	11578	11789	11789	11789.11578	11578	11649	0.2462999999999999999999	0.008620000000000000000000	0.008620000000000000000000	0.008620000000000000000000	0.008620000000000000000000		
31	56.5032	367	328	11908	12032	16.237142857143	370	11942	303741428	11988.57548871	11942	11968	0.2369999999999999999999	0.008680000000000000000000	0.008680000000000000000000	0.008680000000000000000000	0.008680000000000000000000		
32	57.10455	665	665	11789	12424	16.237142857143	12068	12032	12032	12032.12032	12068	12068	0.2865999999999999999999	0.008940000000000000000000	0.008940000000000000000000	0.008940000000000000000000	0.008940000000000000000000		
33	57.10455	665	665	11789	12424	16.237142857143	12068	12032	12032	12032.12032	12068	12068	0.2865999999999999999999	0.008940000000000000000000	0.008940000000000000000000	0.008940000000000000000000	0.008940000000000000000000		
34	58.13873	768	747	11789	12585	18.08451111111111	943	12585	12585	12585.12585	12585	12529	0.1756600000000000000000	0.009250000000000000000000	0.009250000000000000000000	0.009250000000000000000000	0.009250000000000000000000		
35	62.14311	615	571	12585	12875	17.668516429594	239	12875	12875	12875.12875	12875	12839	0.1878000000000000000000	0.009329999999999999999999	0.009329999999999999999999	0.009329999999999999999999	0.009329999999999999999999		
36	62.14311	615	571	12585	12875	17.668516429594	239	12875	12875	12875.12875	12875	12839	0.1878000000000000000000	0.009329999999999999999999	0.009329999999999999999999	0.009329999999999999999999	0.009329999999999999999999		
37	64.00469	12032	11789	12875	12875	14.949795518329	12875	12875	12875	12875.12875	12875	12842	0.2369999999999999999999	0.009470000000000000000000	0.009470000000000000000000	0.009470000000000000000000	0.009470000000000000000000		
38	65.09081	1640	1630	11789	1640	19.904658488555	825	12931	0.035354511	12951	12951	12951	0.1009000000000000000000	0.009550000000000000000000	0.009550000000000000000000	0.009550000000000000000000	0.009550000000000000000000		
39	69.72455	323	315	13317	13516	18.4200000000000000	165	13317	13516	13516.13516	13317	13384	0.138400000000000000000000	0.009580000000000000000000	0.009580000000000000000000	0.009580000000000000000000	0.009580000000000000000000		
40	69.72455	323	315	13317	13516	18.4200000000000000	165	13317	13516	13516.13516	13317	13384	0.138						

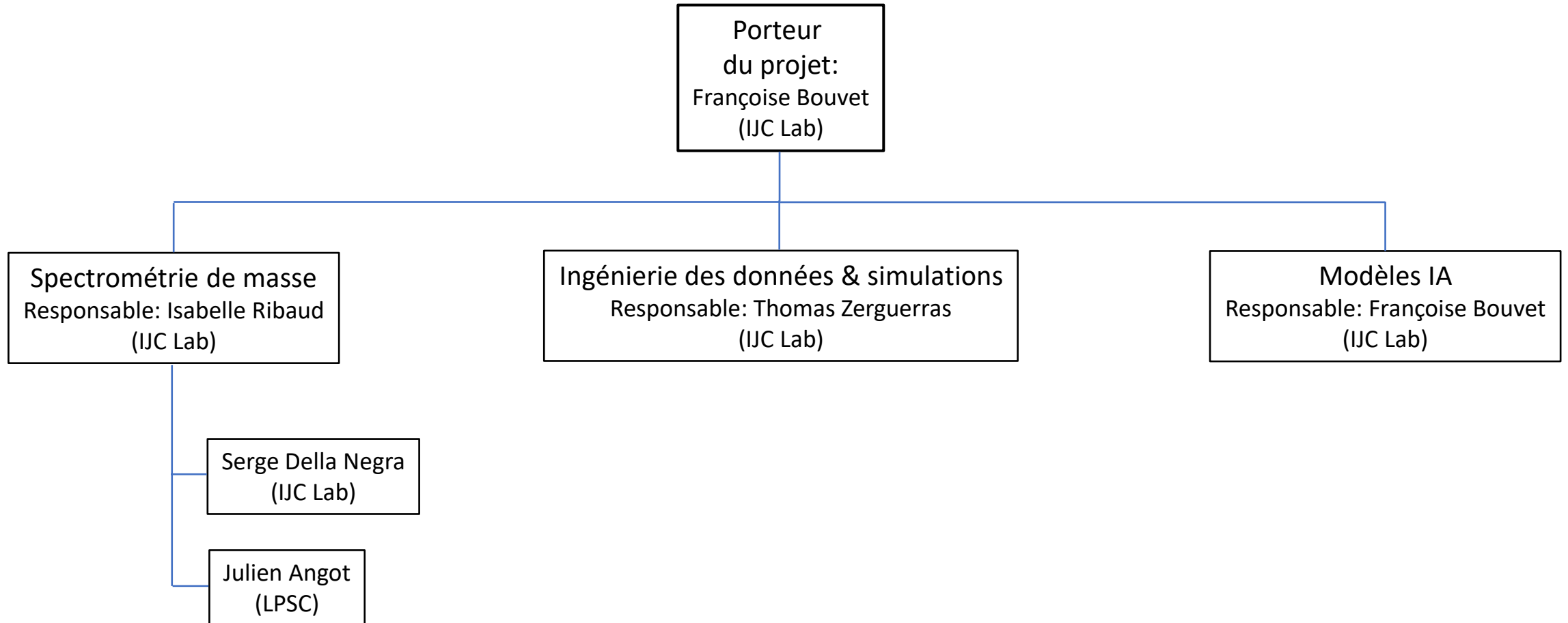


- **Utilisation d'un Perceptron Multi Couche (MLP) pour caractériser les pics**



Option 2 : identification des pics à l'aide d'un réseau de convolution 1D

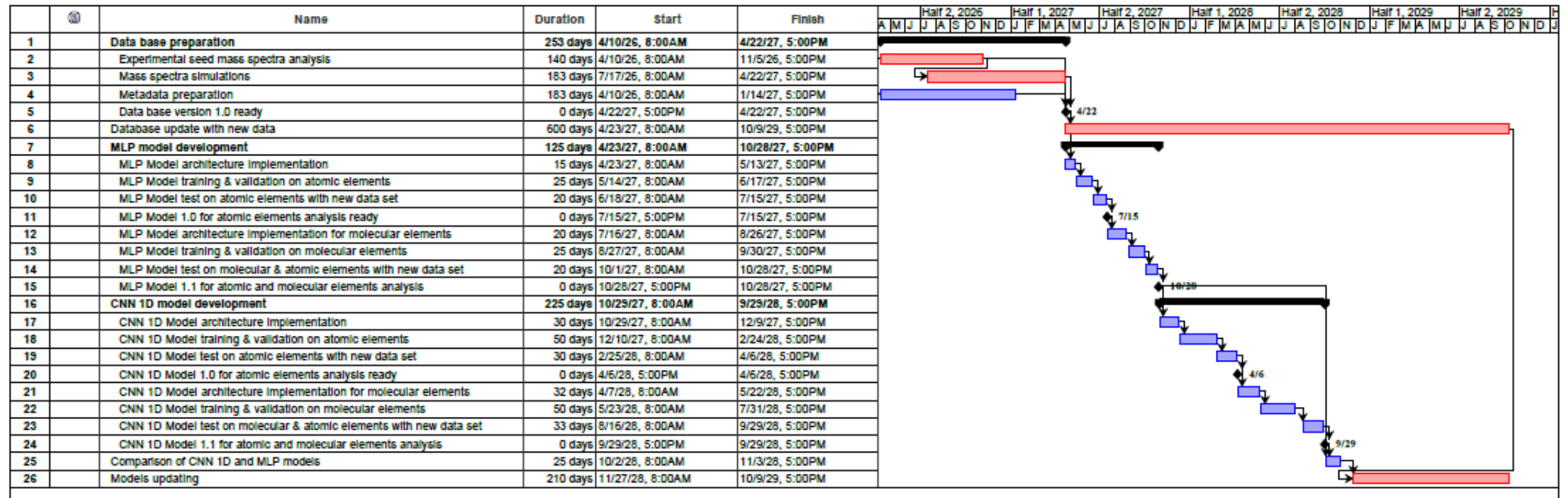






Budget et calendrier (préliminaire)

- **Durée du projet R&T AIMS: 3 ans (2027-2029)**
- **Budget: ~40k€ sur 3 ans dont:**
 - ~27k€ : missions (réunions de travail, workshops, réseaux métiers, GDR, conférences ...)
 - ~13k€ :
 - Capacités de stockage
 - PC avec GPU (pré-test modèles IA)
 - Achat / renouvellement licence (accès à base de données TOF constructeurs)
- **Calendrier**





Conclusion - perspectives

Travail en cours

- Sélection des données expérimentales d'intérêt pour démarrer
- Pré-traitement et préparation des données
- Définition des métadonnées (infrastructure expérimentale, résolution du spectromètre, ...)
- Développement de modules Python pour écriture/lecture de métadonnées (format: imzML)
- Préparation fiche R&T AIMS

A venir

- Développement de la méthode de simulation pour augmenter la quantité de données
- Constitution de la BdD (Andromède, LPSC, Surface Science Spectra et/ou données constructeurs)
- Construction et comparaison des modèles :
 - Extraction des données puis MLP (option 1)
 - Réseau CNN 1D (option 2)
- Adaptation des modèles à des nouvelles données : piège à ion (ISOLTRAP @ CERN), MALDI (Matrix Assisted Laser Desorption Ionization) (I2BC @ Gif-sur-Yvette), ...



Merci de votre attention !